A MODEL FOR THE INTEGRATION OF MICROSATELLITE GENOTYPING WITH PHOTOGRAPHIC IDENTIFICATION OF HUMPBACK WHALES

M.J. ANDERSON, G. HINTEN, D. PATON AND P.R. BAVERSTOCK

Anderson, M.J., Hinten, G., Paton, D. & Baverstock, P.R. 2001 12 31: A model for the integration of microsatellite genotyping with photographic identification of humpback whales. *Memoirs of the Queensland Museum* **47**(2): 451-457. Brisbane. ISSN 0079-8835.

In this study we present a model for the integration of microsatellite genotyping with photographic identification of humpback whales, *Megaptera novaeangliae*, using samples from the east coast of Australia as a case study. A suite of 10 microsatellite markers was selected for this study, based on recommendations made by ANZECC and discussions with other research groups. Seven of the 10 markers were successfully used to genotype 12 sloughed skin samples from humpback whales on their northern migration along the east coast of Australia, resulting in 11 individual whales being identified. Two samples, collected from the same pod of whales, were found to be from one individual, as the genotypes of both samples were identical, while two further samples identified a pair of whales as a possible parent/offspring combination. In order to establish a worldwide database incorporating genetic and photographic identification of humpback whales, results must be standardised between research groups. To overcome potential technical difficulties of standardising results, we recommend that each research group sequence a reference sample or group of reference samples for each locus and that results are reported in repeat number rather than absolute PCR product size. ❐ *Microsatellite genotyping, humpback whale, Megaptera novaeangliae, photo identification.*

Megan Anderson (e-mail: mander15@scu.edu.au), Gavin Hinten & Peter Baverstock, Southern Cross Centre for Whale Research, Southern Cross University, PO Box 157, Lismore 2480; David Paton, Southern Cross Centre for Whale Research, 21 Netherby Rise, Sunrise Beach, Noosa 4567, Australia; 1 August 2001.

Recent studies of humpback whales, *Megaptera novaeangliae*, have employed passive methods such as photographic identification of tail flukes and dorsal fins to examine site fidelity, basic social associations, migratory paths, population estimates and population growth (Isaacs & Dalton, 1992; Gill & Burke, 1999; Garrigue, 2001). Although much knowledge has been derived from photo-ID studies, the technique can be inconsistent and subjective, and susceptible to human error (Corkeron et al., 1999). Furthermore, young humpback whales can undergo extensive colour changes as they grow (Carlson & Mayo, 1990; Valsecchi & Amos, 1996), it is often difficult to approach animals due to behavioural responses and weather conditions, and individuals may lack distinguishing attributes necessary for unambiguous identification (Bain, 1990; Stern et al., 1990; Valsecchi & Amos, 1996).

As a result the Australian and New Zealand Environment and Conservation Council (ANZECC) recommended that genetic analyses be integrated with conventional research methods, such as photographic identification, to address remaining issues concerning humpback whale populations. Microsatellite genotyping is a rapid, accurate and systematic technique, which can provide key insights into humpback whale ecology and evolution. While photo-ID does not lend itself readily to systematic profiling of individuals, it has the advantage of being a simple and obvious method of differentiating between individuals. A digital database incorporating both microsatellite genotyping and photo-ID of humpback whales would combine the advantages of each technique, providing information on population sizes, more detailed social associations, sex identification, mating strategies, stock structure, gene flow and parentage.

Southern Hemisphere humpback whales were classified into six stocks (Groups I-VI) by the International Whaling Commission (IWC) based on their aggregations in Antarctic summer feeding grounds. Genetic analysis of Group IV, V and VI stocks are of particular interest to Australian humpback whale research. Discovery tagging and acoustic analysis of Group IV and V stocks indicate that mixing of these populations is likely to occur (Chittleborough, 1965; Paterson, 1991), as is mixing of Group V and VI stocks (Valsecchi et al., 1997). Genetic differences

within and between these stocks remain unclear and movement patterns of individual whales across jurisdictional boundaries, within and between nations, need further investigation.

This study utilises 10 microsatellite markers to genetically 'fingerprint' 12 humpback whale sloughed skin samples collected during the Cape Byron Whale Research Project 2000. The microsatellite markers were selected as a standard set of genetic markers for humpback whale research in the Southern Hemisphere based on recommendations made by ANZECC (Corkeron et al., 1999) and discussion with other genetics laboratories in the Southern Hemisphere. Our aim was to establish a model for integrating microsatellite genotyping with photo-ID of humpback whales migrating along the east coast of Australia. Such information would provide a basis for establishment of a Southern Hemisphere humpback whale database.

METHODS

SAMPLE COLLECTION. During the Cape Byron Whale Research Project 2000, 91 sloughed skin samples were collected from humpback whales on their northern migration along the east coast of Australia. Where possible, accompanying photographs of the whale's tail fluke and dorsal fins were taken when the skin was collected. Byron Bay was selected for this study due to the close proximity of whales to the mainland, which allowed both land- and sea-based surveys to be conducted. Twelve skin samples were selected for microsatellite genotyping on the basis that each sample could be directly linked to an individual whale by being either the sole animal in a pod, or positively matched to a photo. DNA was extracted from approximately $1cm^2$ of sloughed skin using the Tissue Protocol for the QIAamp DNA Mini Kit (Qiagen) according to the manufacturers instructions, with the exception that extracts were eluted with $2 \times 100\mu l$ of buffer AE, instead of the $2 \times 200\mu l$ recommended.

LABORATORY ANALYSIS. Ten humpback whale microsatellite loci were selected for microsatellite genotyping; EV14, EV21, EV37, EV94, EV96 & EV104 (Valsecchi & Amos, 1996), and GATA28, GATA53, GATA417 & TAA31 (Palsbøll et al., 1997a). The forward primer of each locus was fluorescently labelled so that for each individual, PCR products could be combined for genotyping in two lanes of an automated sequencing gel without products overlapping in colour or expected size range (Table 1).

PCR amplifications were carried out separately for each individual/locus combination before PCR products were combined for gel separation. Locus EV94 was unable to be optimised and was not used for further analysis. PCR reaction mixtures contained: $1 \times$ reaction buffer (Biotech), 0.1mM of each dNTP, $0.1\mu M$ of each of the forward and reverse primers, 0.55 units Taq (Biotech), 2.5mM $MgCl_2$, $4\mu l$ of genomic DNA, and Milli-Q water to a total volume of $20\mu l$. PCR reactions were performed on a PC960G thermal cycler (Corbett Research, Sydney) and run under the following conditions: 1 minute initial denaturation at 92°C, followed by 35 cycles of 10 seconds denaturation at 92°C, 30 seconds annealing at the optimised temperature (Table 1), and 1 minute extension at 75°C followed by a final extension step of 75°C for 5 minutes.

Genotyping of PCR products was conducted on an ABI Prism 310 genetic analyser (Applied Biosystems) using Genescan-500 TAMRA as an internal size standard. Results were displayed using *Genescan* software (Applied Biosystems). Genotypes were scored using *Genotyper* software (Applied Biosystems). Loci GATA417 and TAA31 were unable to be genotyped for the majority of samples and were not used for further analysis.

DATA ANALYSIS. In cases where two identical genotypes were found, the specific probability of identity (POI) for that exact genotype was calculated based on the POI formulae of Paetkau & Strobeck (1994). Due to the low sample size,

TABLE 1. PCR amplification conditions, number of alleles and expected and observed allele size ranges for the 10 microsatellite loci.

Locus	Dye label	Annealing temp. (°C)	Number of alleles	Expected size range (bp)	Observed size range (bp)
EV14	FAM	48	6	125-145	128-142
EV21	FAM	48	6	107-117	107-119
EV37	TET	50	12	190-228	190-218
EV94	TET	-	-	202-222	-
EV96	FAM	48	8	185-213	190-210
EV104	TET	48	3	143-153	141-145
GATA28	TET	48	5	147-191	144-178
GATA53	FAM	48	9	178-210	232-278
GATA417	HEX	50	-	193-293	-
TAA31	TET	48	-	85-121	-

we could not accurately determine allele frequencies and therefore the POI calculations lacked precision, however, the error was small.

RESULTS

DNA from all 12 sloughed skin samples was amplified successfully at 7 loci (EV14, EV21, EV37, EV96, EV104, GATA28 and GATA53). Locus EV94 was unable to be optimised, while GATA417 and TAA31 were optimised successfully for PCR however did not amplify for the majority of samples. All 7 loci that successfully amplified were found to be polymorphic, exhibiting between 3 and 12 alleles (Table 1). The level of allelic diversity detected was similar to that of other studies, despite the comparatively small sample size (Valsecchi & Amos, 1996; Palsbøll et al., 1997a). The expected level of heterozygosity for each locus ranged between 0.55 and 0.98.

Alleles that were potentially unique to this study were detected at 5 loci, EV14, EV21, EV104, GATA28 and GATA53. All samples genotyped at locus GATA53 displayed a marked difference in the size range observed compared to that expected, with as much as a 66 base difference (Table 1). Four other loci displayed alleles outside their expected size ranges, but in each case this was only a difference of one repeat unit.

Samples B73 and B74 displayed identical genotypes at all 7 loci, while no two other samples shared the same genotype at more than 3 loci. These samples were collected in the vicinity of 2 whales migrating together. Samples B3 and B5 were the only other two samples (excluding B73 and B74) which had at least one allele corresponding at all 7 loci genotyped (Table 2). These samples were also collected from within a pod of two whales migrating together.

Accompanying tail fluke photographs were obtained for 6 of the 12 skin samples, with each displaying a large variation in the degree of photo/camera angle, lighting and weather conditions. Photo-IDs were matched alongside their respective microsatellite genotypes for comparative purposes (Fig. 1).

DISCUSSION

For over two decades attempts have been made to photographically identify individual humpback whales around the world. Recent advances in genetic techniques now provide a more informative form of individual identification. Microsatellite genotyping can provide information on contemporary population structure, gene flow, abundances, relatedness and genetic diversity (e.g. McRae & Kovacs, 1994; Richard et al., 1996; Call et al., 1998; Palsbøll et al., 1997b), enhancing the information available from photo-ID research.

In this study, 11 individuals were positively identified using a suite of 7 hypervariable microsatellite loci. The level of variation detected could distinguish all but 2 of the 12 samples from as little as 1 locus (e.g. Fig. 1, GATA53), illustrating the accuracy with which this method can identify individual humpback whales. The two samples (B73 and B74) that could not be distinguished are most likely to be from one whale sampled twice. The specific probability of identity of the exact genotype shared by these two samples is 8.189×10^{-17}; calculated from the frequency of alleles observed, which is an imprecise estimate of the frequency of alleles in

TABLE 2. Genotypes of humpback whale sloughed skin samples for 7 microsatellite loci.

Locus sample	EV14	EV21	EV37	EV96	EV104	GATA28	GATA53
A7	128/ 130	113/ 115	192/ 194	198/ 202	143/ 143	144/ 152	232/ 244
A8	130/ 132	109/ 109	196/ 204	196/ 202	143/ 145	144/ 144	248/ 256
B3	130/ 132	107/ 109	206/ 218	196/ 200	143/ 145	144/ 144	232/ 260
B5	130/ 136	109/ 109	190/ 206	200/ 202	143/ 143	144/ 148	232/ 256
B54	130/ 132	107/ 109	202/ 208	200/ 204	141/ 145	148/ 176	236/ 248
B73	130/ 136	113/ 115	198/ 210	196/ 206	143/ 143	152/ 180	252/ 252
B74	130/ 136	113/ 115	198/ 210	196/ 206	143/ 143	152/ 180	252/ 252
B78	132/ 136	119/ 119	210/ 218	194/ 202	143/ 143	-	-
B92	132/ 132	109/ 111	190/ 192	202/ 202	143/ 145	144/ 152	256/ 264
B97	138/ 142	107/ 115	192/ 194	196/ 196	143/ 143	144/ 144	232/ 248
C2	130/ 130	109/ 109	194/ 206	198/ 200	143/ 143	144/ 152	248/ 252
E10	130/ 130	107/ 115	196/ 212	204/ 210	143/ 143	144/ 144	232/ 276

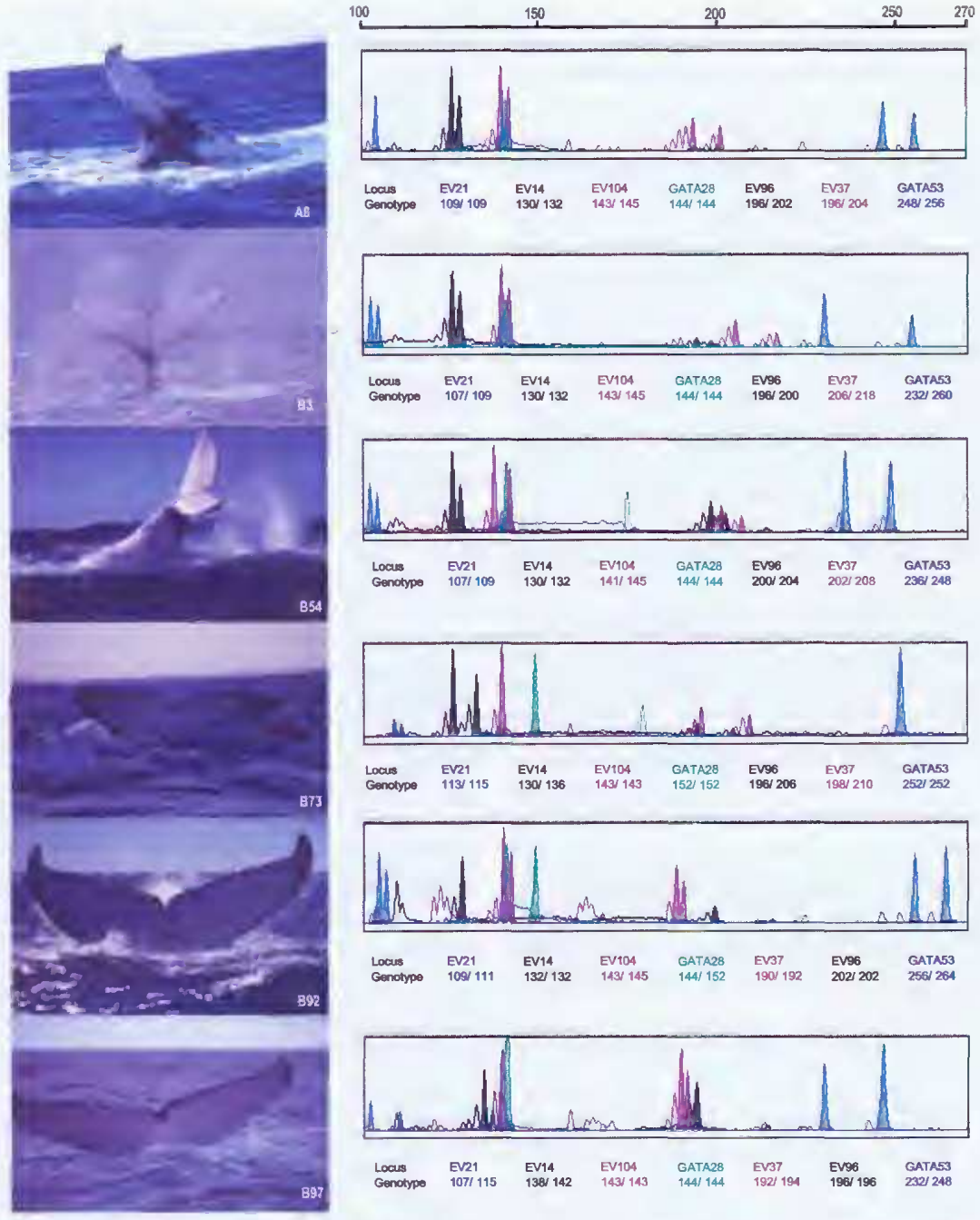

FIG. 1. Individual identifications of humpback whales migrating north past Byron Bay, using photographic identification and microsatellite genotyping. Genotypes are shown for 7 microsatellite loci.

the population, and is therefore only an estimate of the exact probability of identity. Considering the extremely low probability of these two sample genotypes being identical by chance alone and the fact that they were obtained from the one pod of whales, it is likely that these two samples came from the same whale.

This result highlights a potential souree of error when using sloughed skin for microsatellite genotyping purposes. Skin can remain in the water eolumn for up to 20 minutes after bcing dislodged (Corkeron et al., 1999), therefore samples eollected from pods eontaining several whales are less reliable for microsatellite genotyping purposes. Valsecchi et al. (1998) found biopsy darting to be the most effieient method for matching samples to individuals, however this teehnique still cannot guarantee a mateh between a mierosatellite genotype and photo-ID.

Samples B3 and B5 were found to share at least one allele in common at all 7 loci genotyped, suggesting that these two individuals may be related. Other studies have shown that associations between humpbaek whales tend to be non-related (except in the case of mother-ealf pairs) and transient, with few pairs being assoeiated through time (Falcone et al., unpubl. data). Valsecchi et al. (in press) concluded that migrating humpbaek whales did not select their travelling companions based upon relatedness at any stage of the migration. Results of the present study, however, suggest that humpbaek whales may migrate as family units, as both individuals sampled were adults and not a mother-calf pair. Unfortunately, due to the small sample size and the limited number of loci genotyped, the inference of familial relationships based on allele frequencies is not strong. In order to definitively determine potential relationship between individuals, as many as 17 loci may need to be genotyped to minimise the ehance of random matehes (Palsbøll, 1999).

For data to be shared effectively between research groups there are potential technieal errors that need to be addressed, ineluding: non-templated addition of a single adenine base by *Taq* DNA polymerase during PCR (Brownstein et al., 1996; Magnuson et al., 1996); allelic dropout resulting from poor quality template DNA due to degraded or low quantity DNA (Jarne & Lagoda, 1996); null or non-amplifying alleles (Brookfield, 1996; Jarne & Lagoda, 1996); ealibrating PCR product size scoring across hardware; and eonfirmation of amplification of the eorrect locus.

Addition of an adenine base during PCR (+A) can cause problems in allele scoring during genotyping (Magnuson et al., 1996). The frequeney of +A addition ean vary within and between loci, as well as within and between different gel runs, and can be affeeted by different DNA polymerases (Brownstein et al., 1996; Magnuson et al., 1996). Several procedures can be used to overcome this problem. 1) A reference sample or group of referenee samples should be sequeneed for eaeh locus and always included in every PCR and gel. The eorrect product size can then be determined and correct binning boundaries set. 2) Alleles should be recorded as numbers of repeats rather than absolute PCR product size. 3) Different combinations of primer modification and DNA polymerase can be used to either induee 100% +A addition or reduce +A addition to 0%, so that results can be standardised aeeordingly.

When using small quantities of poor quality or degraded DNA, often only one allele of a heterozygous individual is detected (Taberlet & Luikart, 1999). This type of error, called allelic dropout, ereates an artifieial exeess of homozygotes, possibly resulting in departures from Hardy-Weinberg equilibrium. A similar problem is the amplification of null alleles, which occurs when mismatehes in the priming site of one allele cause the failure of that allele to be amplified, again eausing an excess of homozygotes. Allelic dropout and null alleles ean be differentiated as allelic dropout is associated with low quality DNA and therefore ean be detected across loei within an individual, whereas null alleles are assoeiated with a specific locus and can be deteeted aeross individuals within a locus. Another potential problem associated with null alleles, is the use of primers designed for one species to amplify a homologous loeus in another speeies. In sueh instanees more species specifie primers my need to be designed. If reference samples for a locus have been sequeneed and aligned, conserved sequence bloeks ean be identified so that new primers ean be designed for those regions, redueing the risk of null alleles. By reeording results as repeat numbers rather than PCR produet size, allele sizes can be directly eompared between different primer pairs for the same locus.

The use of different hardware for genotyping can result in identieal samples being scored as different sizes. Calibration of hardware within and between researeh groups can be aehieved by sequencing a reference individual or a group of referenee individuals for each loeus, and always ineluding these referenee samples in every PCR and gel. Furthermore, allele seoring can be standardised by reeording data as repeat number rather than PCR product size. A number of loci in this study exhibited what seemed to be extensions

to their known size range, but these may have been the result of incorrect scoring or binning of genotypes, or non-calibration of hardware between laboratories. Hardware calibration and standardised definition of binning boundaries are therefore essential to eliminate potential scoring errors.

When comparing results between research groups it is vital to ensure that the same locus has been amplified in all instances. Sequencing of a reference individual or group of reference individuals will establish whether or not the same locus is being amplified. While such an event may seem unlikely, it did occur in this study. For locus GATA53 we used the primers published in Palsbøll et al. (1997a). Our results showed that the size range differed from the expected by 66 bases. Investigation revealed that one of the published primers was unlikely to be the primer used in that study. Furthermore, when we compared the two primer pair sets on the same individuals, not only did the allele sizes differ, but the relative allele size ranges within individuals also differed. Despite the change in only one primer, it appeared that a different microsatellite locus had been amplified. The only method to test this hypothesis would be to sequence the products for both sets of primers.

This study presents a model for the integration of microsatellite genotyping with photographic identification of humpback whales, using samples from the east coast of Australia as a case study. A standardised digital genetic database would greatly benefit research of humpback whale populations through sharing of results worldwide among research groups and be of immense value for conservation and management purposes. Integration of such a database with current photo-ID databases would enhance the value of each by incorporating the accuracy of microsatellite genotyping with the wealth of photo-ID data available.

ACKNOWLEDGEMENTS

We thank Elena Valsecchi, Scott Baker, and Rob Slade for helpful advice on microsatellite marker selection. We thank Simon Walsh, Wayne Pellow and staff of the NSW National Parks and Wildlife Service and the volunteers of the Cape Byron Whale Research Project 2000 for their assistance in sample collection.

LITERATURE CITED

BAIN, D. 1990. Examining the validity of inferences drawn from photo-identification data, with special reference to studies of the Killer whale (*Orcinus orca*) in British Columbia. Pp. 93-100. In Hammond, P.S., Mizroch, S.A. & Donovan, G.P. (eds) Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. Reports of the International Whaling Commission (Special Issue 12). (Cambridge).

BROWNSTEIN, M.J., CARPTEN, J.D. & SMITH, J.R. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase - primer modifications that facilitate genotyping. Biotechniques 20(6): 1004.

CALL, D.R., HALLETT, J.G., MECH, S.G. & EVANS, M. 1998. Considerations for measuring genetic variation and population structure with multilocus fingerprinting. Molecular Ecology 7: 1337-1346.

CARLSON, C.A. & MAYO, C.A. 1990. Changes in the ventral fluke pattern of the humpback whale (*Megaptera novaeangliae*), and its effect on matching evaluation of its significance to photo-identification research. Pp. 105-111. In Hammond, P.S., Mizroch, S.A. & Donovan, G.P. (eds) Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. Reports of the International Whaling Commission (Special Issue 12). (Cambridge).

CHITTLEBOROUGH, R.G. 1965. Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). Australian Journal of Marine and Freshwater Research 16: 33-128.

CORKERON, P.J., VAN PARJIS, S. & MARSH, H. 1999. Draft proceedings of the Australian and New Zealand Environment Conservation Council (ANZECC) Humpback Whale Research and Conservation Seminar, Brisbane, 1998. (Queensland Department of Environment and Heritage: Brisbane).

GARRIGUE, C. 2001. Characteristics of the New Caledonian humpback whale population. Memoirs of the Queensland Museum 47(2): 539-546.

GILL, P. & BURKE, C. 1999. Whale watching in Australia and New Zealand waters. (New Holland Publishers: Sydney).

ISAACS, R. & DALTON, T. 1992. The Australian guide to whale watching. (Lansdowne Publishing: Sydney).

JARNE, P. & LAGODA, P.J.L. 1996. Microsatellites, from molecules to populations and back [Review]. Trends in Ecology and Evolution 11(10): 424-429.

MAGNUSON, V.L., ALLY, D.S., NYLUND, S.J., KARANJAWALA, Z.E., RAYMAN, J.B., KNAPP, J.I., LOWE, A.L., GHOSH, S. & COLLINS, F.S. 1996. Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase – implications for PCR-based genotyping and cloning. Biotechniques 21(4): 700-709.

McRAE, S.B. & KOVACS, K.M. 1994. Paternity exclusion by DNA fingerprinting, and mate guarding in the hooded seal *Cystophora cristata*. Molecular Ecology 3: 101-107.

PAETKAU, D. & STROBECK, C. 1994 Microsatellite analysis of genetic variation in black bear populations. Molecular Ecology 3: 489-495.

PALSBØLL, P. 1999. Genetic tagging: contemporary molecular ecology. Biological Journal of the Linnean Society 68: 3-22.

PALSBØLL, P., BÉRUBÉ, M., LARSEN, A. & JØRGENSEN, H. 1997a. Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. Molecular Ecology 6: 893-895.

PALSBØLL, P.J., et al. 1997b. Genetic tagging of humpback whales. Nature 388: 767-769.

PATERSON, R.A. 1991. The migration of humpback whales *Megaptera novaeangliae* in east Australian waters. Memoirs of the Queensland Museum 30(2): 333-342.

RICHARD, K.R., WHITEHEAD, H. & WRIGHT, J.M. 1996. Polymorphic microsatellites from sperm whales and their use in the genetic differentiation of individuals from naturally sloughed pieces of skin. Molecular Ecology 5: 313-315.

STERN, S.J., DORSEY, E.M. & CASE, V.L. 1990. Photographic catchability of individually identified minke whale (*Balaenoptera acutorostrata*) of the San Juan Islands, Washington and the Monteray Bay Area, California. Pp. 127-133. In Hammond, P.S., Mizroch, S.A. & Donovan, G.P. (eds) Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. Reports of the International Whaling Commission (Special Issue 12). (Cambridge).

TABERLET, P. & LUIKART, G. 1999. Non-invasive genetic sampling and individual identification. Biological Journal of the Linnean Society 68: 41-55.

VALSECCHI, E. & AMOS, W. 1996. Microsatellite markers for the study of cetacean populations. Molecular Ecology 5: 151-156.

VALSECCHI, E., GLOCKNER-FERRARI, D., FERRARI, M. & AMOS, W. 1998. Molecular analysis of the efficiency of sloughed skin sampling in whale population genetics. Molecular Ecology 7: 1419-1422.

VALSECCHI, E., PALSBØLL, P., HALE, P., GLOCKNER-FERRARI, D., FERRARI, M., CLAPHAM, P., LARSEN, F., SEARS, R., SIGURJONSSON, J., BROWN, M., CORKERON, P. & AMOS, W. 1997. Microsatellite genetic distances between oceanic populations of the humpback whale (*Megaptera novaeangliae*). Molecular Biology and Evolution. 14(4): 355-362.

VALSECCHI, E., HALE, P., CORKERON, P. & AMOS, W. (in press). Social structure in migrating humpback whales (*Megaptera novaeangliae*). Molecular Ecology.